Math 170S

Lecture Notes Section 7.3 [*,†]

Confidence intervals for proportions

Instructor: Swee Hong Chan

NOTE: Materials that appear in the textbook but do not appear in the lecture notes might still be tested. Please send me an email if you find typos.

---

*Version date: Saturday 31st October, 2020, 23:23.

†This notes is based on Hanbaek Lyu's and Liza Rebrova's notes from the previous quarter, and I would like to thank them for their generosity. "*Nanos gigantum humeris insidentes* (I am but a dwarf standing on the shoulders of giants)".

1 Confidence intervals for proportions: Example

In a certain political campaign, one candidate conducted a poll for which 185 out of 351 voters favor this candidate. This candidate then calculates the percentage of people who voted for them, which is

$$\frac{185}{351} \approx 0.527.$$

This is higher than 50%, and the candidate now feels confident about winning. Is the candidate's confidence justified?

2 CI for proportions: Settings

- **Object:** Y is a **Bernoulli** random variables with **unknown parameter** p.

- **Input:**

  - Random samples $y_1, \ldots, y_n$ for Y. Note that each y_i is either 0 or 1.

  - Confidence constant $1 - \alpha$

- **Output:** The value ε that allows us to say

  "p is contained in the interval $\left[\,\overline{\mathrm{y}} - \varepsilon, \overline{\mathrm{y}} + \varepsilon\right]$ with confidence (approximately) 1-α."

The interval $\left[\,\overline{\mathrm{y}} - \varepsilon, \overline{\mathrm{y}} + \varepsilon\right]$ is the **confidence interval** for p. This interval is centered at $\overline{\mathrm{y}}$, and the length of the interval is 2ε.

3 The value ε

Theorem 1. *In this case, the value ε is given by*

$$\varepsilon = z_{\alpha/2}\sqrt{\frac{\overline{\mathrm{y}}(1-\overline{\mathrm{y}})}{n}},$$

where $z_{\alpha/2}$ is the real number such that

$$P[N(0,1) \geq z_{\alpha/2}] \;=\; \alpha/2.$$

The value $z_{\alpha/2}$ can be computed from the Table V in Appendix B. □

4 Answer for Example

Suppose that confidence constant $1-\alpha$ is 0.95. Then

$$\bar{y} = \frac{185}{351} = 0.527;$$
$$\varepsilon = z_{\alpha/2}\sqrt{\frac{\bar{y}(1-\bar{y})}{n}} = (1.96)\sqrt{\frac{(0.527)(0.473)}{351}} \approx 0.052.$$

The 95% confidence interval is equal to

$$\big[(0.527) - 0.052, (0.527) + 0.052\big] = \big[0.475, 0.579\big].$$

Hence there is some possibility that p is less than 0.5, so the candidate should be very careful when campaigning as we **cannot say that we are 95% sure that they will win the election**.

Remark 2. Note that textbook has two other formulas ((7.3.4) and (7.3.5) in the textbook) for estimating ε. We will only focus on the formula provided in the lecture notes for now.

5 Settings: One-sided intervals

- **Object:** Y is a **Bernoulli** random variables with **unknown parameter** p.
- **Input:**
  - Random samples $y_1, \ldots, y_n$ for Y. Note that each y_i is either 0 or 1.
  - Confidence constant $1 - \alpha$
- **Output:** The value ε that allows us to say

  "p is contained in the interval $\left[0, \bar{\mathrm{y}} + \varepsilon\right]$ with confidence 1-α,"

  or

  "p is contained in the interval $\left[\bar{\mathrm{y}} - \varepsilon, 1\right]$ with confidence 1-α."

6 Value of ε: One-sided

Theorem 3. *In this case, the value ε is given by*

$$\varepsilon = z_\alpha \sqrt{\frac{\overline{\mathrm{y}}(1-\overline{\mathrm{y}})}{n}},$$

where z_α is the real number such that

$$P[N(0,1) \geq z_\alpha] \; = \; \alpha.$$

The value z_α can be computed from the Table V in Appendix B. □

7 Answer for Example: One-sided

We again take the confidence constant $1 - \alpha$ to be 0.95. In an election, we only care if the candidate gets more than 50% of the vote, so we want

"p is contained in the interval $\left[\,\overline{\mathrm{y}} - \varepsilon, 1\right]$ with confidence (approximately) 1-α,"

Then we have

$$\overline{\mathrm{y}} = \frac{185}{351} = 0.527;$$
$$\varepsilon = z_\alpha \sqrt{\frac{\overline{\mathrm{y}}(1 - \overline{\mathrm{y}})}{n}} = (1.645)\sqrt{\frac{(0.527)(0.473)}{351}} \approx 0.044.$$

The confidence interval is then equal to

$$\big[(0.527) - 0.044, 1\big] \;=\; \big[0.483, 1\big].$$

Again we see that p can be less than 0.5 in this confidence interval, so we **cannot say that we are 95% sure that they will win the election**.

Let's try to calculate the winning probability of our candidate, i.e., we want to find α so that

"p is contained in the interval $\left[0.5, 1\right]$ with confidence (approximately) 1-α."

This implies that (BT)

$$\begin{aligned}
\overline{\mathrm{y}} - \varepsilon &= 0.5 \\
z_\alpha \sqrt{\frac{\overline{\mathrm{y}}(1-\overline{\mathrm{y}})}{n}} &= \overline{\mathrm{y}} - 0.5 \\
z_\alpha &= \frac{(\overline{\mathrm{y}} - 0.5)\sqrt{n}}{\sqrt{\overline{\mathrm{y}}(1-\overline{\mathrm{y}})}} \\
z_\alpha &= \frac{(0.527 - 0.5)\sqrt{351}}{\sqrt{(0.527)(1-0.527)}} \\
z_\alpha &\approx 1.01.
\end{aligned}$$

By Table V, $1 - \alpha$ is approximately 0.8438. So the candidate's chance of winning is 84.38%.

8 Example: Comparing two proportions

Two disinfectants were tested for their ability to remove coronavirus from a dry surface.

- The first detergent is successful on 63 out of 91 trials;
- The second detergent is successful on 42 out of 79 trials.

Can we say that one detergent is stronger than the other confidently?

9 Two proportions: Settings

- **Object:** Y_1, Y_2 are **independent Bernoulli** random variables with **unknown parameter** p_1, p_2.
- **Input:**
  - Sample mean $\overline{y}_1$ for Y_1 from n_1 many samples, and sample mean $\overline{y}_2$ for Y_2 from n_2 many samples;
  - Confidence constant $1 - \alpha$.
- **Output:** The value ε that allows us to say

"$p_1 - p_2$ is contained in the interval $\left[(\overline{y}_1 - \overline{y}_2) - \varepsilon, (\overline{y}_1 - \overline{y}_2) + \varepsilon\right]$ with confidence (approximately) 1-α."

10 The value ε: Two proportions

Theorem 4. *In this case, the value ε is given by*

$$\varepsilon = z_{\alpha/2}\sqrt{\frac{\overline{\mathrm{y}}_1(1-\overline{\mathrm{y}}_1)}{n_1} + \frac{\overline{\mathrm{y}}_2(1-\overline{\mathrm{y}}_2)}{n_2}},$$

where $z_{\alpha/2}$ is the real number such that

$$P[N(0,1) \geq z_{\alpha/2}] \;=\; \alpha/2.$$

The value $z_{\alpha/2}$ can be computed from the Table V in Appendix B. □

11 Answer for Example: Two proportions

Suppose that confidence constant $1-\alpha$ is 0.9. Then (BT)

$$\begin{aligned}\overline{y}_1 - \overline{y}_2 =& \frac{63}{91} - \frac{42}{79} = 0.16;\\ \varepsilon =& z_{\alpha/2}\sqrt{\frac{\overline{y}_1(1-\overline{y}_1)}{n_1} + \frac{\overline{y}_2(1-\overline{y}_2)}{n_2}}\\ =& (1.645)\sqrt{\frac{(0.692)(0.308)}{91} + \frac{(0.532)(0.468)}{79}} \approx 0.12192.\end{aligned}$$

The confidence interval is then equal to

$$\left[(0.16)-0.12192, (0.16)+0.12192\right] = \left[0.03808, 0.28192\right].$$

Because the interval lies entirely to the right of zero, we can say that the first disinfectant is better than the second one with at least 90% confidence.